Monte Carlo Sampling Methods Homework 2

Steven (Zihan) Zhang

September 28, 2022

0.1 Exercise 16

Following the logistic introduced in Example 5, the CDF is:

$$F(x) = \sqrt{x}, x \in [0, 1] \tag{1}$$

given that $\pi(x) = \frac{1}{2\sqrt{x}}, x \in [0, 1]$. Thus, for $U \sim \mathcal{U}[0, 1]$, we could generate the distribution $Y = F^{-1}(U)$:

$$F^{-1}(u) = u^2, u \in [0, 1] \tag{2}$$

$N = 10^7$. Figure 1 describes the QQ plot.

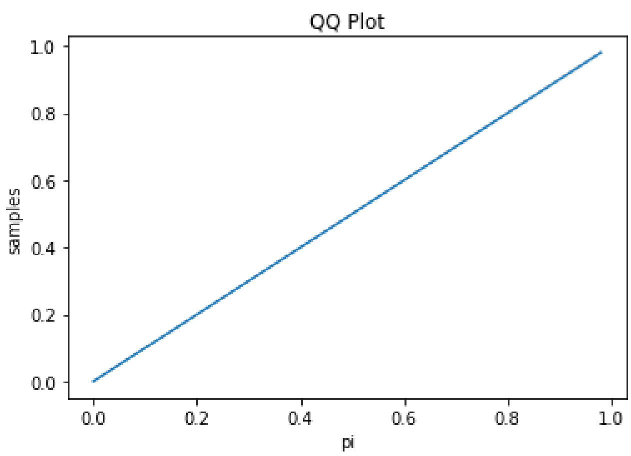

Figure 1: Comparison of samples to π.

0.2 Exercise 18

Similar to Example 6 of Box-Muller, we could generate uniform distribution on disk. Suppose $u_1, u_2 \sim \mathcal{U}[0, 1]$ that are independent and identically distributed random variables (i.i.d.). Thus:

$$\begin{aligned} x &= \sqrt{u_1}\cos(2\pi u_2) \\ y &= \sqrt{u_1}\sin(2\pi u_2) \end{aligned} \tag{3}$$

The Jacobian matrix is:

$$A = \begin{bmatrix} \frac{1}{2\sqrt{u_1}}\cos(2\pi u_2) & -2\pi\sqrt{u_1}\sin(2\pi u_2) \\ \frac{1}{2\sqrt{u_1}}\sin(2\pi u_2) & 2\pi\sqrt{u_1}\cos(2\pi u_2) \end{bmatrix} \tag{4}$$

The determinant is π. Since u_1, u_2 are generated from the uniform distribution and $u_1 = x^2 + y^2$:

$$\begin{aligned} \pi(u_1, u_2) &= 1|_{0 \le u_1, u_2 \le 1} \\ \pi(x, y) &= \frac{1}{\pi}|_{0 \le x^2 + y^2 \le 1} \end{aligned} \tag{5}$$

which is the density for x and y. Figure 2 describes the histogram. The cost time for single sample is $5.9127 \times 10^{-8}s$.

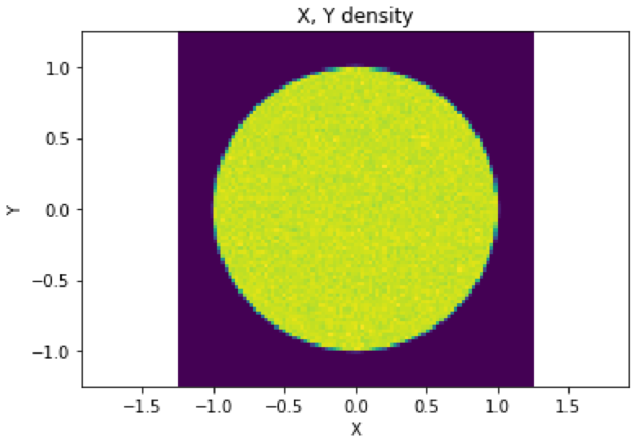

Figure 2: 2-D histogram that describes generating a uniformly distributed sample on the unit disk given two independent samples from $\mathcal{U}(0,1)$.

0.3 Exercise 19

We could sample from $\tilde{\pi} = \frac{1}{4}|_{-1 \leq x,y \leq 1}$. Our target is $\pi = \frac{1}{Z}|_{0 \leq x^2+y^2 \leq 1}$. We could set $K = 4/Z$ since $1 \leq Z \leq 4$. Here $\pi/K\tilde{\pi}$ is the indicator function $1|_{0 \leq x^2+y^2 \leq 1}$. The expected number of $\mathcal{U}(0,1)$ variables required per sample from the unit disk is ≈ 1.2731. The cost time (expected wall clock time) for a single sample is $6.7824 \times 10^{-6}s$.

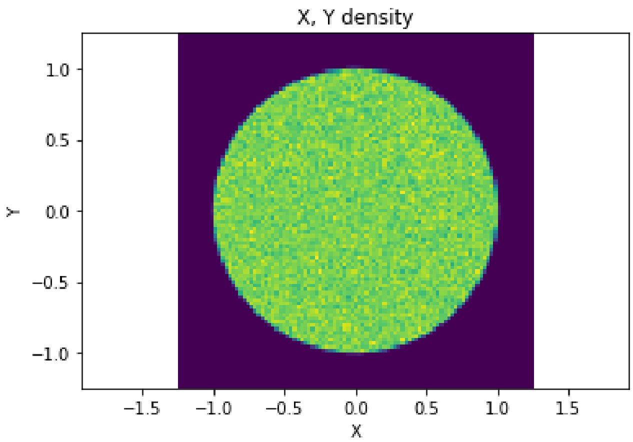

Figure 3: 2-D histogram.

0.4 Exercise 20

We know that $\pi = \mathcal{N}(0,1), \tilde{\pi} = \mathcal{N}(m, \sigma^2)$. Also, the f function is $f(y) = 1_{y \geq 2}$. Thus:

$$\frac{\pi}{\tilde{\pi}} = \sigma e^{-x^2/2+(x-m)^2/2\sigma^2} \tag{6}$$

By Page 34 in Chapter 3, our estimator would be:

$$\tilde{f}_N = \frac{1}{N}\sum_{k=1}^{N} f(y^{(k)})\frac{\pi(y^{(k)})}{\tilde{\pi}(y^{(k)})} \tag{7}$$

where $y^{(k)} \sim \tilde{\pi}$. We run the experiments with different parameters in a relatively fine grid $\Delta = 0.5$. To yield the best estimators, we have $m = 2.5$ and $\sigma = 0.5$.

0.5 Exercise 21

We have:

$$\begin{aligned} p(x) &= e^{-|x|^3} \\ q(x) &= \tilde{\pi}(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ Z_q &= 1 \end{aligned} \tag{8}$$

Thus:

$$\frac{1}{N} \sum_{k=1}^{N} \frac{p(Y^{(k)})}{q(Y^{(k)})} \to \frac{Z_p}{Z_q} = Z_p \tag{9}$$

The calculating result with 10^6 samples is 1.7866, which is very close to the numerical integration result (also presented in the log output).

0.6 Exercise 22

Very similar to Exercise 20. The only difference is we change the importance sampling estimator $\tilde{f}_N / \tilde{1}_N$ instead of $\tilde{f}_N$. Thus, we would replace the important sampling estimator as:

$$\sum_i \frac{f(y^{(i)}) \frac{\pi(y^{(i)})}{\tilde{\pi}(y^{(i)})}}{\sum_j \frac{\pi(y^{(j)})}{\tilde{\pi}(y^{(j)})}} \tag{10}$$

where $y^{(i)} \sim \tilde{\pi}$. The optimal m and σ would be 1.5 and 1.5. We prefer the estimator mentioned in Exercise 20 since the variance is smaller and the computation speed is faster, meaning the performance is better.